If a Human Can See It, So Should Your System: Reliability Requirements for Machine Vision Components

Boyue Caroline Hu boyue@cs.toronto.edu
Lina Marsso lina.marsso@utoronto.ca
Krzysztof Czarnecki kczarnec@gsd.uwaterloo.ca
Rick Salay rsalay@gsd.uwaterloo.ca
Huakun Shen huakun.shen@mail.utoronto.ca
Marsha Chechik chechik@cs.toronto.edu

Overview

Machine Vision Components (MVCs) are deployed in safety-critical systems, where undesired behaviors can lead to fatal accidents. **Towards safe MVCs, one needs to define what it means for an MVC to be correct and then check its correctness prior to system deployment.**

MVC reliability against scene changes:
The performance of an MVC should remain reliably unaffected by scene changes that can occur in real-world scenarios.

Using human performance as a baseline, we define MVC reliability as:

If the changes do not affect humans, they shouldn't affect MVC either.

Motivating Example

Consider an autonomous driving scenario during winter, when frost can develop.

Image credit: https://www.wired.com

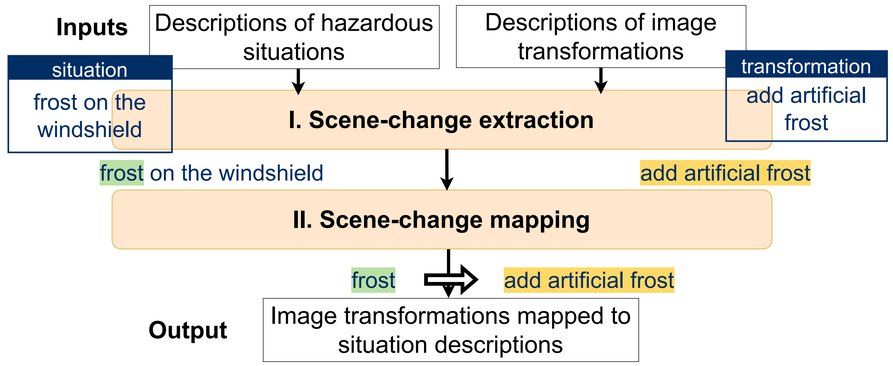
To check MVC reliability:
1. Select transformations simulating frost
2. Obtain reliability requirements
3. Check requirements satisfaction

Implementations

Paper

Transformation Selection

We defined a method for systematic and automatic selection of image transformations simulating scene changes described in hazardous situation specifications.



Frost on the windshield can be simulated with this transformation that adds artificial frost.

Reliability Requirements

We cannot require MVCs to remain reliable subject to arbitrary changes in the environment.

original image
$\Delta_V = 0$

minimal changes
$\Delta_V = 0.005$

reasonable changes
$\Delta_V = 0.71$

unreasonable changes
$\Delta_V = 0.96$
not recognizable by humans

For a classification MVC, given the artificial frost, we specified two types of reliability requirements:

1. (*Correctness-preservation*) The recognition accuracy of an MVC should not decrease if the visual change in the images is within the range $\Delta_V <= 0.84$
2. (*Prediction-preservation*) The percentage of labels an MVC can preserve after adding frost should not decrease if visual change in the images is within the range $\Delta_V <= 0.91$

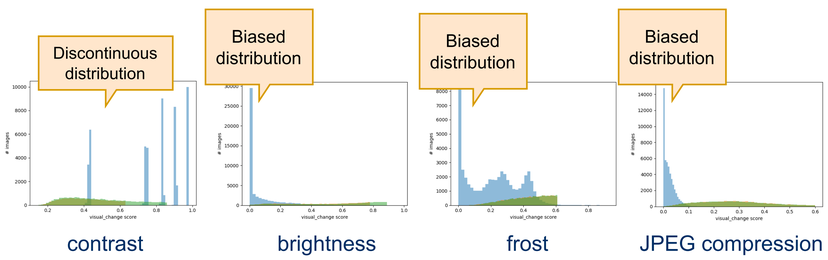
- Δ_V is a measure of visual change in images, value range: [0, 1]
- Threshold values are estimated from human performance data

Correctness-preservation VS Prediction-preservation

Correctness-preservation	Prediction-preservation
• Checks the **correctness of decisions** after transformation	• Checks the **preservation of decisions** after transformation
• Requires **ground truth** which is costly to obtain	• Can be checked on **unlabeled images** which are easier to obtain
• If an MVC satisfies only the correctness-preservation requirement, it may **correctly recognize different objects** before and after transformation	• If only the prediction-preservation requirement is satisfied, the MVC might **preserve incorrect decisions** and change correct ones

Evaluation

Comparing tests in state-of-the-art benchmark CIFAR-10-C with our tests for checking requirements satisfaction.



Also, by checking our requirements, we can capture reliability gaps missed by state-of-the-art benchmarks. E.g., for JPEG compression

CIFAR-10-c leaderboard model name	Rank on CIFAR-10-c	Rank on satisfying correctness-preservation	Rank on satisfying prediction-preservation
RLAT	1	5	1
RLATAugMix NoJSD	2	2	7

CIFAR-10-c only ranks with accuracy. Our results show that high accuracy does not guarantee reliability against changes that do not affect humans.